Jun 19th, 9:00 AM - 10:20 AM

The Virtual Learning Commons (VLC): Enabling Sharing and Innovation for Flexible, Responsive Solutions

Deana D. Pennington
University of Texas at El Paso, ddpennington@utep.edu

Alida Gándara
University of Texas at El Paso, agandara1@utep.edu

Nick Del Rio
University of Texas at El Paso, ndel2@utep.edu

Omar Ochoa
University of Texas at El Paso, omar@miners.utep.edu

International Environmental Modelling and Software Society (iEMSs)
7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA
Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2014-proceedings

The Virtual Learning Commons (VLC): Enabling Sharing and Innovation for Flexible, Responsive Solutions

Deana D. Pennington[1], Aida Gándara[1], Nick Del Rio[1], Omar Ochoa[1]

University of Texas at El Paso: ddpennington@utep.edu, agandara1@utep.edu, ndel2@utep.edu, omar@miners.utep.edu

Abstract: A key challenge in modeling of complex natural systems is identifying and integrating models and tools that are relevant, but were not developed for the targeted purpose. Modeling resources are proliferating, along with emerging technologies designed to enable discovery and integration of data and models. The goal of the Virtual Learning Commons (VLC) is to improve awareness and understanding of emerging methodologies and technologies, facilitate individual and group evaluation of these for innovative uses, and trace the impact of innovations across research communities. The VLC is funded by the U.S. National Science Foundation, and will be released in 2014. The VLC is a Web-based social bookmarking site with support for semantic analysis of content. It makes use of Web 2.0 (Social Web) and Web 3.0 (Semantic Web) approaches to enable discovery of potentially relevant model and tool innovations, and networked learning about their function. It is founded on well-developed models of technology adoption and diffusion of innovation. Expected outcomes from the VLC project are:

- Improved dissemination of emerging models and tools across disciplines and institutions using Semantic Web approaches;
- Better support for community learning using social and networked learning approaches; and
- New methods for capturing the provenance (trace) of innovations from creators, to interested researchers, to users, based on the new provenance standard PROV-O.

The VLC project is significant because across the globe, decision makers are calling for better translation of research products into impacts. The VLC provides a 21st century approach towards innovation, enabling better sharing and reuse of models and tools in a wider variety of research contexts.

Keywords: semantic web; social media; networked learning; diffusion of innovations; technology adoption.

1. INTRODUCTION

Integrated environmental modeling (IEM) applications are resource intensive, involving a wide variety of data, models, and modeling strategies (Laniak et al., 2013). A key challenge is discovering, understanding, and leveraging existing resources (Laniak et al., 2013). Not only has the IEM community generated a plethora of resources, rapid developments in information technologies have resulted in a deluge of new capabilities (Pennington, 2011). It is challenging to even be aware that resources exist that are potentially relevant to a particular application. The best documentation of a new resource is often in the research literature of the community that produced the resource, which is not likely to be read or understood by a potential adopter from a different domain.

At the same time, creators of new technologies and methodologies designed to aid scientists often struggle to have impact (Bietz et al. 2013). Programs such as the U. S. National Science Foundation EarthCube purposefully partner computer, information, and cognitive scientists with earth and environmental scientists for the purpose of exploring community needs. Yet the challenge of efficiently sharing knowledge and resources across disciplines remains, and it seems likely that there will continue to be a deluge of new methods and data, some of which will be vastly underused simply because it is difficult for potential users to learn about them.

Two workshops were conducted at a U. S. Long Term Ecological Research (LTER) Network All Scientists Meeting in September 2009 to generate a community vision of a Web-based system that could support more effective learning about application of technologies. Forty-eight scientists, information managers, and other disciplinary experts attended. Participants showed high interest in community-generated content in four different areas: 1) technical explanations of approaches and tools; 2) awareness of social and organizational context, e.g. knowing who created the approach, who has used the approach, and who might be interested in collaborating around the approach; 3) explanations of how a specific technology is actually used -- the work flow in real settings; and 4) understanding of the scientific content in the context where the technology is being used.

This article describes the outcome from the LTER workshops, the Virtual Learning Commons (VLC). VLC is a Semantic Web-based social bookmarking tool for discovering, organizing, discussing, analyzing, and visualizing research products and resources. The goal of the VLC is to support three kinds of users: 1) researchers who are trying to make sense of unfamiliar technical choices available to them (tool seekers); 2) researchers who have technical products that they are disseminating (tool providers); and 3) researchers who study the diffusion of innovation within and across communities (innovation analysts). The VLC aims to connect tool seekers and providers through a networked learning environment (Goodyear et al., 2004) that can provide data for innovation analysts. In Section 2 we provide an overview of the VLC. Section 3 provides technical details regarding the use of semantic and social approaches in the VLC. Section 4 is discussion.

2. THE VIRTUAL LEARNING COMMONS (VLC) APPROACH

The VLC is being designed based on questions that drive our three user types. These include: What tools are available for accomplishing x? Who is working on those tools, from what institutions? Where has it been applied on the ground? When was the tool developed? How does the tool work? How has it impacted research within and beyond the discipline of its origination?

The VLC leverages Semantic Web approaches (Berners-Lee 2001; W3C http://www.w3.org/; Figure 1) to create, analyze, and visualize connections between people and resources. Semantic Web is a rapidly evolving set of technologies for encoding meaningful descriptions (semantics) into Web resources in machine-readable forms. Standard computer languages and ontologies[1] are used to describe Web resources, which structures Web content in ways that allow computers to reason about the content. The VLC employs an ontology created for the purpose of describing research impacts (the VLC Research Influence Model (RIM) ontology[2]) as well as other commonly used ontologies such as the Friend-of-a-Friend (FOAF) ontology (http://www.foaf-project.org/). The VLC operates on Web links that are identified and organized by a human user. For this task, we have created a plug-in to the Zotero (https://www.zotero.org/) research assistant (Figure 1). Zotero provides functionality for collecting, organizing, and citing research sources, especially documents. A key feature of Zotero is its ability to automatically extract metadata, e.g. citation information, from web pages using bibliographic standards and customized translators, allowing the user to add citations to their personal library with a single click. Our plugin to Zotero, called VLCWeb, provides the same one-click functionality to extract semantic content.

The VLC and VLCWeb make use of *resources* and *collections*. A resource is any content that is available to the user through the VLC/VLCWeb. Resources have unique, independent existence in the VLC. A specific resource (URL) is only in the VLC database as a single instance. Any comments, ratings, or discussion about that resource are attached to that instance, and appear in any context where the resource appears. A given resource may be added to one to many *collections*. A VLC

[1] An ontology is a formal specification of a conceptualization of a domain, including its entities, their properties, and relationships.
[2] RIM Ontology namespace: http://ontology.cybershare.utep.edu/RIM.owl

collection is a container that a user creates to group related resources. This could be, for example, a collection of resources that describe tools that a researcher is considering using on a particular project. A collection could also be links to competing projects that a researcher wants to monitor.

A special type of collection is a *project,* which is a collection for which specific project-related semantics are defined. VLC projects have titles, investigators, start and end dates, all of which are specified semantically. When resources are added to a project collection, each resource inherits the semantics of the project. Additional semantics can be added manually, but this ensures that many resources added in the VLCWeb will have at least a partial semantic description -- since many research-related resources of interest are associated with a research project.

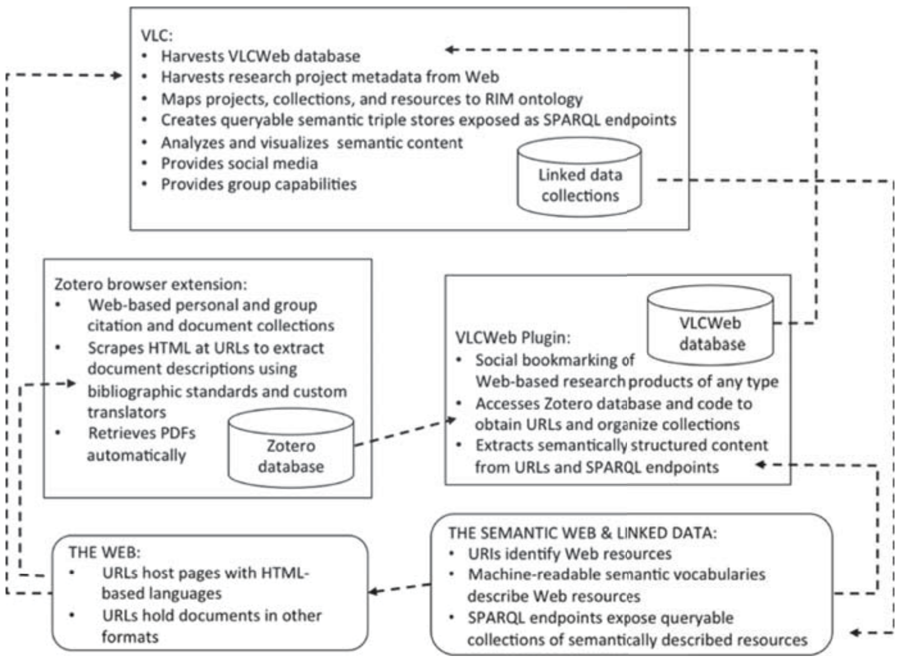

Figure 1. Key components of the VLC that facilitate the capture and analysis of Web and Semantic Web information.

3. LINKED DATA, SEMANTICS, AND PROVENANCE IN THE VLC

3.1 VLCWeb: Research Products as Linked Data

The VLC collects both structured (semantically encoded) and unstructured information about Web resources using various techniques. Structured information on the Web is encoded using the Resource Description Framework (RDF; http://www.w3.org/RDF/), and may or may not be exposed as a *SPARQL endpoint* -- a store of RDF that can be queried (http://www.w3.org/TR/rdf-sparql-query/). When a structured Web resource is linked in the VLC, the VLC automatically extracts all of the RDF. This approach does not limit the semantic vocabulary of Web resources to the forms recognized by the VLC; rather, it enhances the structure of information integrated into VLC collections to include existing semantic vocabularies already on the Semantic Web. In addition to preserving existing semantics, unique identifiers of resources on the Semantic Web, i.e., a resource's URI, are preserved. We expect that RDF availability will grow with time as semantic approaches become more widespread, and this will be our primary mechanism for obtaining semantic content. However, for now little semantic content is available; therefore other (semi-)automated processes are being incorporated into the VLC. One extraction technique in the VLC leverages Zotero's functionality for automated sensing and information extraction from web pages. Web information is harvested using Zotero's functionality and is automatically transformed into VLC-compliant structured information. Project semantics have also been obtained through programmatically traversing websites and scraping information. For example, the VLC has harvested a total of 29776 funded research projects from the NSF website and automatically structured that information into RDF. 210 of those were selected as a testbed, had corresponding VLCWeb project collections created, and were grouped into nine collections of related projects. Web resources are being added to these projects in order to demonstrate semantic analysis and visualization capabilities. When the VLC exposes its collections on the Web, they are exposed as enhanced linked data (http://linkeddata.org/) through a SPARQL endpoint enabling other semantic applications to query and reuse VLC content.

3.2 VLC-Enabled Exploration and Discovery

Through a variety of relevant visualizations, the VLC can enable understanding of the information found in a collection. Currently, the VLC displays four main views of a collection: a synopsis, a location view, a collaborator view and a tag cloud (Figure 2). The synopsis displays a table of aggregate information such as the number of people, resources, and locations available in a collection. By selecting any of aggregate values, the user can explore views in detail. For example, when building a collection named Sensor Networks, scientists could identify NSF-funded projects that are either developing or using sensor networks, and add these to the collection. They could add Web resources to the projects, such as links to presentations or training materials. They could add links to datasets that were generated by sensors; the metadata for these presumably contain coordinates for the location of the samples. Visualizing the Sensor Networks collection would display locations where the datasets were collected, emphasizing the kinds of environments in which particular sensors are being used (for example, aquatic vs. terrestrial). Alternatively, a collaborator view would enable the user to see who is conducting research on sensor networks, and who is collaborating with whom. All VLC views are provided in a similar fashion - users see a summary of the view, can select components to visualize a detailed view, and can switch between views to achieve different perspectives on the collection content.

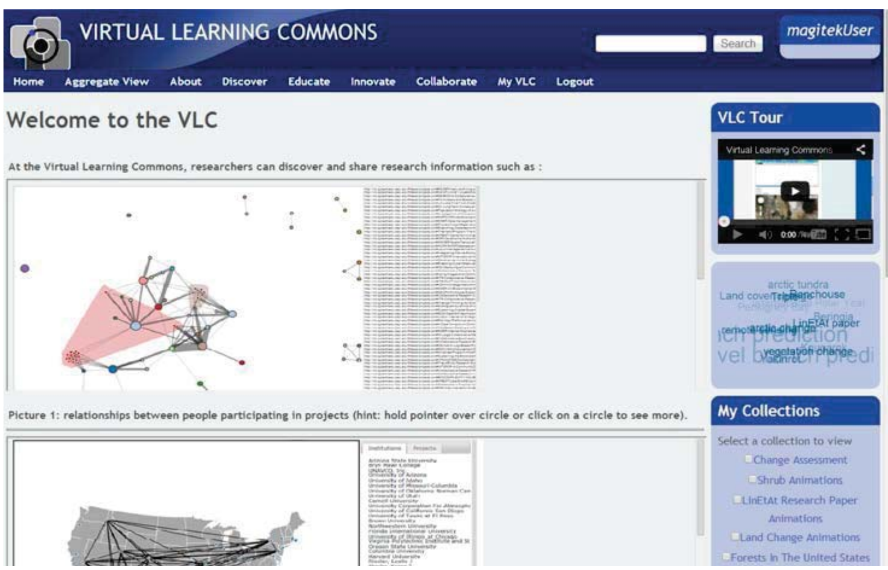

Figure 2. Screenshot of the VLC home page, showing the collaborator view of a collection, tag cloud, and a portion of the spatial view.

VLC supports human understanding about a collection through faceted searches in the detailed views using the semantic properties found in a collection. A faceted search enables users to select properties, i.e., facets, to browse and apply filters. To support this, the VLC determines human understandable identifiers from the machine-readable encoding, making faceted searches using semantic-based vocabulary more user-friendly. By enabling users to explore resources in different views and leveraging faceted searches to gain further understanding, the VLC aims to support more efficient discovery of potentially relevant existing research tools than can be attained on the distributed Web.

3.3 SOCIAL MEDIA AND NETWORKED LEARNING IN THE VLC

VLC collections meet the first need and most of the second from the LTER requirements workshops described in the introduction. Determining who might be interested in collaborating, interacting with current users of a technology, and understanding the science requires social interaction. For this reason, social media approaches (comments, ratings, etc.) are being integrated into the VLC to support its function as an informal *networked learning environment*. Networked learning may be defined as learning in which information technologies are used to *promote connections* between one learner and other learners; between learners and tutors; and between a learning community and its learning resources (Goodyear et al., 2004, p. 1). In the case of the VLC, the learning community is the research community. “Tutors” are anyone in the research community who can explain a methodology, technique, or tool to someone unfamiliar with it. We refer to tutors as mentors, since learning in this setting occurs through informal exchanges of knowledge between colleagues who act as mentors rather than formal educators. Learning resources are the variety of materials on the Web that could be used to support that effort recognized by the VLC, along with social media within the VLC.

The VLC provides a networked learning *environment*, which is a prerequisite for, but insufficient to produce networked learning. Networked learning by definition requires interaction between humans mediated by digital resources. While certainly an individual can learn in important ways in the VLC simply by interacting with digital resources, it is our contention that innovation will occur most effectively when the resources mediate interactions between humans - either colleagues in the same domain working together to understand a new technology, or researchers from different domains. The VLC seeks to promote the *development of connections* between tool seekers, mentors, and tool providers leveraging digital resources as mediators. While the VLC is being designed to support certain kinds of interactions, any networked learning environment is selectively appropriated by participants who adapt it to their own contexts (Jones, 2011). Hence, a key design approach has been to construct some (hopefully) useful social functionality, then evaluate how people use the VLC in order to make decisions about future social features.

In addition to typical social media approaches (comments and ratings), we have identified a few features that we believe will catalyze discussion within the VLC. Users will be able to create groups who can have their own private collections. Although the resources in any VLC collection are public, discussion can be private. Users have the ability to take a snapshot of a specific view and share that snapshot with a specific individual or group. The snapshot will become a part of the VLC social content to which comments and ratings can be attached.

Studies of innovation in general and social websites in particular indicate that it is critical to target specific communities as early adopters, tailor the new product to those communities, and have active advocates within each community (Moore, 1991). Towards that end, we have target communities (including LTER) whose interests we are familiar with, who have already been engaged. Our initial release will include resources that are relevant to those communities. Key people within those communities have agreed to act as advocates. However, we seek to construct a lightweight model of peer production (Haythornthwaithe, 2009) that encourages participation by as many people as possible, with small contributions by any individual that do not require coordination with others.

3.4 Tracing Innovation in the VLC: Provenance of Ideas

Rogers’ (2003) theory of diffusion of innovations has been used widely in scholarly literature to describe how individuals come to adopt a new technology. His five-stage model consists of: (a) first exposure, (b) interest, (c) decision weighing, (d) trial runs, and (e) commitment to continued use. Case studies of technology adoption within the information technology community have identified many factors influencing this process. The interactive exploration and analysis provided by the VLC is expected to promote discovery and awareness of potentially relevant technologies, having a positive impact on the early stages of innovation. From other perspectives - those of tool providers or innovation analysts - VLC provides an opportunity to assess how a project's research products are being used by other projects and conduct collective analysis of these interactions across all projects. The RIM ontology (Figure 3) was created specifically for this.

The RIM ontology and its supporting system models research efforts by relating projects and people. The system provides mechanisms to assert information about projects, i.e. title, PI and Co-PI, funding agency, etc., and relationships such as *preceding results*, and *new results*. The *preceding results* relationship describes the contributing resources that are being used by this project coming from some other project source. The *new result* relationship describes the resource that this project is generating that can be used by some other project. These properties are then used to infer *influence* among projects and people. In turn these inferences are used to generate D3 (Bostock et al., 2011) visualizations by different perspectives: person-to-person, person-to-project, and project-to-project.

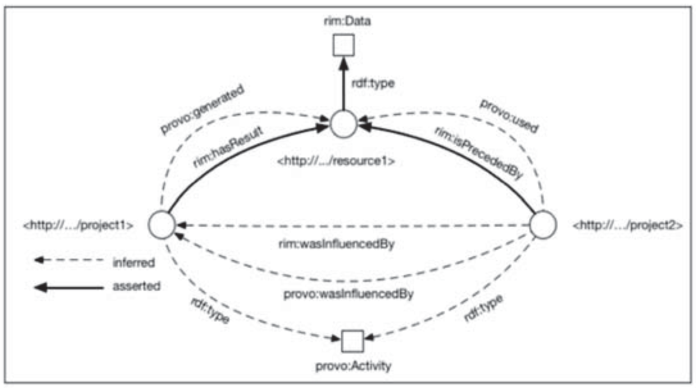

Figure 3. The RIM ontology supports rules for inferring Prov-o relationships from RIM assertions.

RIM is encoded as an OWL ontology that extends existing project-centric ontologies such as the ARPFO (http://vocab.ox.ac.uk/projectfunding) and FRAPO (http://www.essepuntato.it/lode/http://purl.org/cerif/frapo) with influence specific modeling concerns. In particular, the RIM ontology extends prior models by introducing a set of properties such as rim:hasResult and rim:isPrecededBy that specialize the relationship between projects and associated resources. Figure 3 presents how these properties are used in practice to associate a rim:Data resource with two different research projects. The dataset resource in this example was a *result* of project 1 while that same resource *preceded* the work supported by project 2. It is important to note that the asserted relationships are sourced by evidence gathered from the social interactions of users through the provided VLC networking commons. The inferred relationships are derived by the application of the RIM ontology rule set as well as the rule set provided by PROV-O. In the figure, the RIM rule set derived the rim:wasInfluencedBy relationship between <http://.../project2> and <http://.../project1> as well as the specific PROV-O relationships since rim:hasResult and rim:isPrecededBy are subproperties provo:generated and provo:used respectively. The domains of these PROV-O properties, in turn, trigger the two projects to be classified as PROV-O activities. Providing this provenance based perspective to the project influence landscape allows easier processing of RIM by third-party querying and visualization systems that are equipped to handle PROV-O. In this case, the project influence space can be queried and visualized in a similar manner as scientific provenance traces.

With the inclusion of RIM as a backend-reasoning system, VLC users can discover and link resources as provided by the VLC social functions while remaining unaware of the underlying knowledge base being constructed. When users make certain associations such as tagging resources in terms of resultant or preceded work, RIM will assert the appropriate statements in the knowledge base, possibly triggering the reasoner to infer additional relationships.

4. DISCUSSION

While research resources can be discovered on the Web through traditional search engines, a search within existing search engines typically returns millions of results for any general search terms entered. The results mix research on that topic with everything else that is available, and entry-level materials suitable for a potential adopter are difficult to find. The VLC aims to identify a web within the Web that supports the research community by limiting results to those that are related to research projects or have been identified by researchers as appropriate. These high level materials -- such as research project education, outreach, and training materials -- are critical to the ability of potential adopters to enter and progress along stages of technology adoption. At the same time, links in the VLC are not limited to these materials. Any Web resource that a VLC user believes to be useful can be added. This includes resources about technologies that are well known to some but not to others, not just products from recent research. Innovation occurs in different times in different communities. Breakthroughs in a given community can occur through the new application of technologies that previously emerged in other fields. The VLC, using the RIM model, is a first step for collecting data to investigate these complex processes of innovation. Furthermore, semantically enriched data and models are becoming more common (Villa et al., 2009). Because of its automatic extraction of all semantic information available for a resource, in the future the VLC will be able to leverage these semantics to recommend scientific resources that are compatible with each other and in a particular context. The VLC could be linked with knowledge-based modeling systems that are able to integrate semantically described data and models (Villa et al., 2009), providing users a mechanism for discovering and adding relevant components to such a system.

To date, work on the VLC has focused on development of semantic approaches, supporting individual interaction with digital resources. Social approaches are fairly limited. Nevertheless, we see social networking and networked learning as a key need that the VLC must support. A known barrier to any tool supporting groupwork is the need for there to be significant community usage before it becomes useful for any individual (Grudin, 1994). The success of the VLC will depend on engaging a few key communities whose needs for this kind of support may invoke a willingness to invest time in contributing resources.

The IEM community is positioned to become one of these key communities. Laniak et al. (2013, p. 7) note there is a “wide range of approaches for designing, executing, and documenting applications, making it difficult to understand, review, and reuse applications.” They call for a community recognized, standardized process and set of practices for documenting applications, including not just the software used (data, models, etc.) but also the context and expertise used during the application. This is very similar to the needs of the LTER community noted above, including 1) technical explanations; 2) social and organizational context; 3) the workflow in real settings; and 4) the scientific context of the application. The VLC could potentially facilitate the collection of information required to understand a particular IEM application and collaboratively evaluate its potential for reuse in a new context. More generally, the VLC aims to provide infrastructure to support communities of practice (Lave and Wenger, 1991), a recognized need in the IEM community (Laniak et al., 2013). Laniak et al. (2013, p. 17) call for web-based infrastructure that “links the global community and facilitates sharing of knowledge and products, provides environment for collaboration and education, and promotes open and transdisciplinary culture and practices.” This accurately describes the goals of the VLC.

5. CONCLUSIONS

The VLC is a semantic-based, social bookmarking Web application designed to support the research community. Its goal is to provide easy mechanisms for users to discover and learn about technologies relevant to their work, either individually or as communities. Over the long term, we envision that the VLC could become a standard dissemination method for funded research projects whereby investigators add links to relevant high-level resources for others to discover more easily than through a Web search. We envision the VLC serving as an online marketplace for research tools, where buyers and sellers (seekers and providers) can interact in useful ways that are not available through other means.

6. ACKNOWLEDGMENTS

This work has been funded by the National Science Foundation under grant numbers OCI-1135525 for the CI-Team Diffusion project: The Virtual Learning Commons, and HRD 1242122 for the Cyber-ShARE Center of Excellence. The authors gratefully acknowledge a variety of students at the University of Texas at El Paso who have provided input and feedback on the VLC. In particular, we acknowledge Ariel Garcia, Javier Garza, Matthew Giandoni, Edgar Padilla, Joaquin Reyna, and Matthew Rister.

7. REFERENCES

Berners-Lee, T., Hendler, J., & Lassila, O. 2001. The Semantic Web. *Scientific American*, May 11, 2001.

Bietz, M., Paine, D., & Lee, C. 2013. The work of developing cyberinfrastructure middleware projects. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 1527–1538). New York, NY, USA: ACM.

Bostock, M., Ogievetsky, V., & Heer, J. 2011. D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2301–2309.

Goodyear, P., Banks, S., Hodgson, V., & McConnell, D. 2004. Research on networked learning: An overview. In P. Dillenbourg, M. Baker, C. Bereiter, Y. Engeström, G. Fischer, H. U. Hoppe, … D. McConnell (Eds.), *Advances in Research on Networked Learning* (pp. 1–9). Springer Netherlands.

Grudin, J. 1994. Groupware and social dynamics: Eight challenges for developers. *Communications of the ACM*, *37*(1), 93–105.

Haythornthwaite, C. 2009. Crowds and communities: Light and heavyweight models of peer production. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on* (pp. 1–10). IEEE.

Jones, C. 2011. Networked learning environments. *Physical and Virtual Learning Spaces in Higher Education: Concepts for the Modern Learning Environment. Hershey, PA: IGI Global*, 102–118.

Laniak, G. F., Olchin, G., Goodall, J., Voinov, A., Hill, M., Glynn, P., … Hughes, A. 2013. Integrated environmental modeling: A vision and roadmap for the future. *Environmental Modelling & Software*, *39*, 3–23.

Lave, J., & Wenger, E. (1991). *Situated learning: legitimate peripheral participation*. Cambridge University Press.

Moore, G. 1991. *Crossing the Chasm*. New York: Harvard Business.

Pennington, D. 2011. Collaborative, cross-disciplinary learning and co-emergent innovation in informatics teams. *International Journal of Earth System Informatics*, *4*(2), 55–68.

Rogers, E. M. 2003. *Diffusion of Innovation* (Vol. 5th). New York: Free Press.

Sun, H., & Zhang, P. 2006. The role of moderating factors in user technology acceptance. *International Journal of Human-Computer Studies*, *64*, 53–78.

Villa, F., Athanasiadis, I. N., & Rizzoli, A. E. 2009. Modelling with knowledge: A review of emerging semantic approaches to environmental modelling. *Environmental Modelling & Software*, *24*, 577–587.